UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Retention of Exemplar-Specific Information in Learning of Real-World High-Dimensional Categories: Evidence from Modeling of Old-New Item Recognition

Permalink
https://escholarship.org/uc/item/9kv4b9v0

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Nosofsky, Robert
Meagher, Brian

Publication Date
2022

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Retention of Exemplar-Specific Information in Learning of Real-World High-Dimensional Categories: Evidence from Modeling of Old-New Item Recognition

Robert M. Nosofsky (nosofsky@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. Tenth Street
Bloomington, IN 47405 USA

Brian J. Meagher (bmeagher@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. Tenth Street
Bloomington, IN 47405 USA

Abstract

Participants learned to classify a set of rock images into geologically-defined science categories. We then investigated the nature of their category-based memory representations by collecting old-new recognition data in a subsequent transfer phase. An exemplar model provided better qualitative accounts of the old-new recognition data than did a prototype or clustering model. However, to account for the variability in recognition probabilities among the old training items themselves, a hybrid-similarity exemplar model was needed that took account of distinctive features present in the items. The study is among the first to use computational models for making detailed quantitative predictions of old-new recognition probabilities for individual items embedded in complex, high-dimensional similarity spaces.

Keywords: categorization; old-new recognition; high-dimensional similarity spaces; computational models

An important question in cognitive science concerns the relation between the fundamental cognitive processes of categorization and old-new recognition memory. Although some theorists hypothesize that these processes are governed by separate representational systems (e.g., Knowlton & Squire, 1993), others have proposed single-representation-system accounts (e.g., Nosofsky, 1988, 1991; Nosofsky et al., 2011; Nosofsky & Zaki, 1998). Assuming a single-representational system, patterns of old-new recognition data have the potential to provide important evidence bearing on alternative models of categorization. Whereas various models may be difficult to distinguish based on examination of categorization data alone, the joint modeling of categorization and old-new recognition often provides highly diagnostic constraints.

In past work, Nosofsky (1988, 1991) illustrated an exemplar-based approach to the joint modeling of categorization and recognition. According to the exemplar model, people represent categories by storing individual exemplars of the categories in memory. Classification decisions are based on the summed similarity of a test item to the exemplars of a target category *relative* to its summed similarity to contrast categories. By comparison, old-new recognition decisions are based on the *absolute* summed similarity of the test item to all the exemplars of all the categories. This absolute-summed similarity provides a measure of global activation or "familiarity", with greater degrees of familiarity leading to higher "old" recognition probabilities (e.g., Gillund & Shiffrin, 1984; Osth & Dennis, in press). Because different decision rules are involved (a relative- vs. an absolute-summed-similarity rule), the exemplar model is able to account in quantitative detail for both the categorization and recognition of individual items based on their locations in a multidimensional similarity space – even when there are dissociations in performance across the two tasks (for extensive past illustrations, see, e.g., Nosofsky, 1988, 1991, 2017).

However, past work involving exemplar-based modeling of individual-item old-new recognition is limited in at least two respects. First, in previous applications, the domain of inquiry involved use of artificial stimuli varying across relatively few perceptual dimensions. In recent work, Nosofsky and colleagues have extended the exemplar model to account for categorization of real-world, high-dimensional stimuli, namely rock types as formalized in the geologic sciences (e.g., Nosofsky, Sanders, & McDaniel, 2018a,b; Nosofsky, Meagher, & Kumar, 2020, in press). However, there has been no work testing the model on its ability to account for old-new recognition of these real-world high-dimensional stimuli. One purpose of the present research was to begin to fill that gap and test the exemplar model of recognition (and other competing models) in this real-world, high-dimensional category domain. A second purpose of the present research was to address why some old items are easier to recognize than others. Most of the past success of the exemplar model in predicting individual-item old-new recognition has involved the prediction of false-alarm rates associated with novel lures. In general, as lures become more similar to previously experienced target items, their false alarm rates increase, a pattern that is captured naturally by the exemplar model. By contrast, for a variety of reasons, there has been relatively little work testing the ability of the exemplar model to account for differences in hit rates associated with the old target items themselves.

In the present research, we conducted an experiment in which participants learned to classify a large set of rock images into the geologically-defined broad divisions of igneous, metamorphic, and sedimentary rocks. Following a classification-learning phase, there was a test phase in which

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. ©2022 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

participants classified both the old training items as well as new transfer items from the categories. In addition, participants judged whether each test item was old or new. Among the new transfer items were a set of photoshopped rock images that we constructed to be highly similar to specific old items from the training set. We refer to this set of photoshopped images as the "high-similarity-neighbor" (HSN) transfer items; and to the training items from which they were constructed as the "parent" training items.

Our main focus in the present article is on the old-new recognition data. Our initial goal was to use the recognition data to obtain converging evidence bearing on the nature of the category representations that participants developed for learning the classifications. According to prototype models, people learn categories by averaging across the training instances of the categories and representing each category in terms of its central tendency. Because the training items, standard transfer items, and HSNs all tend to be roughly equally distant to the central tendency of their category, the prototype model predicts extremely poor ability to make old-new discriminations among these item types (e.g., Homa et al., 2019; Hu and Nosofsky, 2021). According to clustering models (e.g., Anderson, 1991; Love, Medin, & Gureckis, 2004), people represent individual categories in terms of multiple clusters, with each cluster summarized by its own prototype. Items that are highly similar to one another join together into merged clusters. Clustering models can predict that old items will be recognized as old with higher probability than the standard novel transfer items, because they are more likely to strongly activate the specific clusters that have been formed during the classification-learning phase. However, as will be seen, they are likely to have difficulty in predicting above-chance old-new discrimination for the old training items vs. the HSN transfer items: Because those item types are extremely similar to one another, they are likely to activate the same clusters to roughly the same degree. Finally, exemplar models have the capability of predicting above-chance old-new discrimination ability for all three item types: Because the relation between similarity and distance in psychological space is highly nonlinear, the exact match between a training item and its exemplar representation in memory can lead to a higher overall familiarity signal than is yielded by the near-match of an HSN transfer item (see Modeling section for details).

To preview, we will see that the broad qualitative pattern of results in our old-new recognition data favors the predictions from the exemplar model. However, we will see that even that model fails dramatically to account for the extensive variability in old-recognition probabilities observed within the class of old-training items themselves. We then take steps to extend the model to begin to address this challenge. Although the extended model makes movement in the right direction, future research will be needed to provide a more satisfactory account of the complete set of old-training-item recognition results, a challenge for essentially all models in the field (cf., Bainbridge, 2019).

Method

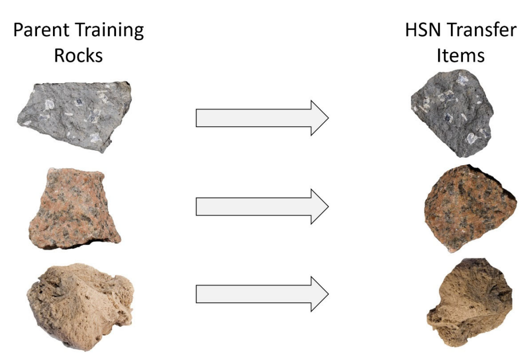
Participants

The participants were 105 Amazon Mechanical Turk workers. We removed from analysis data from 23 participants who performed at near-chance levels on the primary classification task itself, leaving 82 participants.

Stimuli and Apparatus

The stimulus set was composed of 480 standard rock images described in depth in previous articles (Nosofsky et al., 2018a,b; Sanders & Nosofsky, 2020), with an additional 60 photoshopped rock images. The 480 standard rock images consisted of 16 images from each of 30 major subtypes of rocks that are commonly taught in introductory geoscience classes. There were 10 subtypes from each of the three broad divisions of igneous, metamorphic, and sedimentary rocks. Roughly speaking, each subtype can be viewed as a "basic-level" category in the world of geology; whereas the three broad divisions exist at a superordinate level. For each subtype, six randomly chosen items served as training items that participants learned to classify during an initial training phase. Four of these six training items were "non-parent" training items (a total of 120 non-parent training items) and two were "parent" training items (a total of 60 parent training items). The remaining 10 items from each subtype were standard transfer items that were presented only during the test phase (a total of 300 standard transfer items). The 60 photoshopped rock images were constructed to be "high-similarity neighbor" (HSN) transfer items. In particular, each HSN was highly similar to one of the 60 specific parent-training rocks, with 2 HSNs per each of the 30 subtype categories. Examples are provided in Figure 1. See Nosofsky et al (in press) for other illustrations of the HSNs and for the detailed procedures for constructing them.

Figure 1: Examples of Parent Training Rocks and High-Similarity-Neighbor Transfer Items



Procedure

The experiment started with a training phase: each of the 180 total training examples (120 non-parents and 60 parents) was shown once per block in a random order across 3 blocks for a total of 540 training trials. On each trial of the training phase, a rock image was presented and the participant attempted to classify it into one of the 3 broad categories. Immediate feedback was provided on each trial informing the participant of the correct category response. Following training there was an immediate test phase. For each of the 30 subtypes, in addition to the 6 old training examples, participants were presented with the 10 novel standard transfer items and the 2 HSN transfer items (for a total 180 old training items, 300 standard transfer items, and 60 HSN transfer items). The 540 test rocks were presented in a random order for each participant. On each trial, participants first judged the broad category division to which the rock belonged, and then judged whether the rock was "old" (an item experienced during the training phase) or "new". No corrective feedback was provided during the test phase.

Results

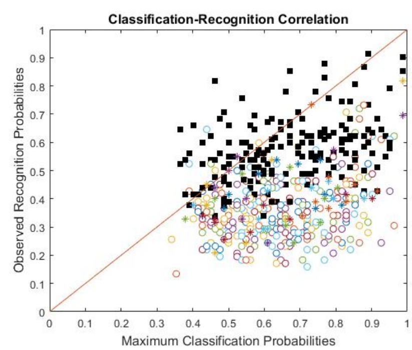
The mean "old" recognition probabilities for the four main item types are reported in Table 1. A one-way repeated-measures ANOVA revealed a main effect of item type, $F(3, 243) = 113.3$, $MSE = 0.009$, $p < .001$. Pairwise t-tests showed that the parent training items were judged "old" with significantly higher probability ($M=.567$) than the HSNs ($M=.409$), $t(82) = 9.59$, $p < .0001$; and the HSNs were judged "old" with significantly higher probability than the standard transfer items ($M=.356$), $t(82) = 6.73$, $p < .0001$.

Table 1: Mean "Old" Recognition Probabilities for the Different Item Types.

Model	Non-Parent Training	Parent Training	Standard Transfer	HSN-Transfer
Observed	0.575	0.567	0.356	0.409
Prototype	0.434	0.434	0.434	0.434
Rational	0.448	0.449	0.424	0.436
Exemplar	0.556	0.556	0.355	0.454
Hybrid-Sim Exemplar	0.564	0.574	0.351	0.446

A central question concerns the ability of the alternative old-new recognition models to capture these overall differences in endorsement probabilities across the item types. Prior to launching into the formal modeling analyses, however, we first consider the extent to which the patterns of categorization and old-new recognition judgments may be related. One simple hypothesis is that categorization and recognition are guided by essentially the same underlying cognitive mechanisms and decision rules. According to such an hypothesis, if an observer is highly confident that an item belongs to a category then the observer will also have a high probability of endorsing the item as "old". As a proxy for the "categorization confidence" associated with each individual item, we use the maximum probability with which each item was classified by observers into each of the broad-division categories. For example, if observers classify item i as "igneous" with probability near one, then classification confidence for that item is high; whereas if observers classify the item into each of the three categories with roughly equal probability, then classification confidence is low. In Figure 2 we plot the "old" recognition probabilities associated with each item against this classification-confidence measure. The solid points in the scatterplot correspond to the old training items; the open circles to the standard transfer items; and the asterisks to the HSN transfer items. It is apparent from inspection that the relation between the old-recognition probabilities and classification confidence is weak: the correlation between the two measures is only $r = .29$. Nevertheless, we will provide preliminary model-based evidence below that classification confidence may have provided a significant residual contribution to observers' old-new recognition judgments.

Figure 2: Correlation Between Categorization "Confidence" and Old-Recognition Judgments



Testing the Formal Models of Individual-Item Old-New Recognition

In this section we test different members of the class of global-activation models on their ability to account for the individual-item old-new recognition data. We start by testing baseline versions of prototype, clustering, and exemplar models. All three classes of models make reference to a high-dimensional MDS solution for the rock images derived from extensive similarity-scaling and dimension-ratings studies reported in previous articles (Nosofsky, Sanders, Meagher, & Douglas, 2018, 2020; Sanders & Nosofsky, 2020). We emphasize that beyond accounting in quantitative detail for the similarity-ratings data, the dimensions derived from this previous MDS scaling work all had natural psychological interpretations.

For all three classes of models, each item i is presumed to give rise to a global activation of memory, or "familiarity", denoted F_i. The probability that item i is judged to be "old" is then given by $P(Old|i) = F_i^{\gamma} / (F_i^{\gamma} + k)$, where γ is a power-transform response-scaling parameter and k is a response-criterion parameter. The models differ only in terms of how the global-activation F_i is computed.

According to the prototype model, each category is represented in terms of the central tendency of each category distribution, computed by averaging across the values of the training items along each of the MDS dimensions. Let x_{im} denote the value of item i on dimension m, and let P_{Jm} denote the value of the prototype of category J on dimension m. The Euclidean distance of item i to prototype J is given by $d_{iJ} = [\sum|x_{im}-P_{Jm}|^2]^{1/2}$; and the similarity of item i to Prototype J is given by $s_{iJ} = \exp(-\kappa \cdot d_{iJ})$, where κ is an overall sensitivity parameter for translating distance to similarity (Shepard, 1987). The familiarity of item i is then computed by summing the similarity of i to the three prototypes, $F_i = \sum s_{iJ}$. The prototype model uses three free parameters: γ, *k*, and κ.

The exemplar model is similar to the prototype model, except that instead of summing the similarity of item i to category prototypes, one sums the similarity of item i to each of the individual training exemplars. The distance of item i to exemplar j is given by $d_{ij} = [\sum|x_{im}- x_{jm}|^2]^{1/2}$; the similarity of item i to exemplar j is given by $s_{ij} = \exp(-\kappa \cdot d_{ij})$; and the familiarity for exemplar i is given by $F_i = \sum s_{ij}$. The exemplar model uses the same free parameters as the prototype model: γ, *k*, and κ.

Our representative from the class of clustering models is Anderson's (1991) rational model. Our implementation of the rational model follows closely the presentation provided by Anderson (1991, pp. 411-414), but a detailed listing of the equations would exceed the length limits of this article. As described in our introduction, the general idea is that the training exemplars are grouped into clusters during the category-training phase, and each cluster is summarized in terms of its own prototype and standard deviations along the component dimensions. There are two free parameters that govern which clusters are formed: a coupling parameter *c*, and a category-label salience parameter α. When *c* is set at a high value, the model tends to group many stimuli together into large clusters, and vice-versa when *c* is set at a low value. At time of test, one computes the probability that a test item i belongs to each of the clusters J, pclus(J|i), as well as the probability that it belongs to a "new" cluster that has not yet been formed, pclus(new|i). Our measure of "familiarity" for the rational model is found by summing the probability that item i belongs to each of the old clusters, $F_i = \sum$ pclus(J|i). The baseline version of the rational model uses four free parameters: γ, *k*, *c* and α. We should clarify several aspects of our fitting of the rational model. First, the clusters that are formed if the goal is to learn to categorize may be different from those that are formed if the goal is to recognize. However, to provide the model with flexibility, we searched for the free parameters in the model that optimized its fits to the old-new recognition data without the constraint of categorization. Second, the clusters formed by the rational model will vary depending on the precise sequence of stimuli with which it is trained. Therefore, fitting the model requires use of computer simulation. In the present case, for any given set of candidate parameter values, our fits were based on averages computed across 10,000 simulations, with a different random training sequence used in each simulation.

We fitted the baseline prototype, rational, and exemplar models to the individual-item old-new recognition data by conducting computer searches for the values of the free parameters that minimized the Bayesian Information Criterion, BIC = -2ln(L) + Pln(N), where L is the maximum-likelihood of the data given the model, P is the number of free parameters in the model, and N is the number of observations in the data set. The model that minimizes BIC is viewed as providing the most parsimonious account of the data. Using multiple random starting configurations, we used the Hooke and Jeeves (1961) search algorithm to locate the best-fitting parameters. As will be seen, our conclusions based on BIC are corroborated by salient qualitative patterns in our model-fitting results.

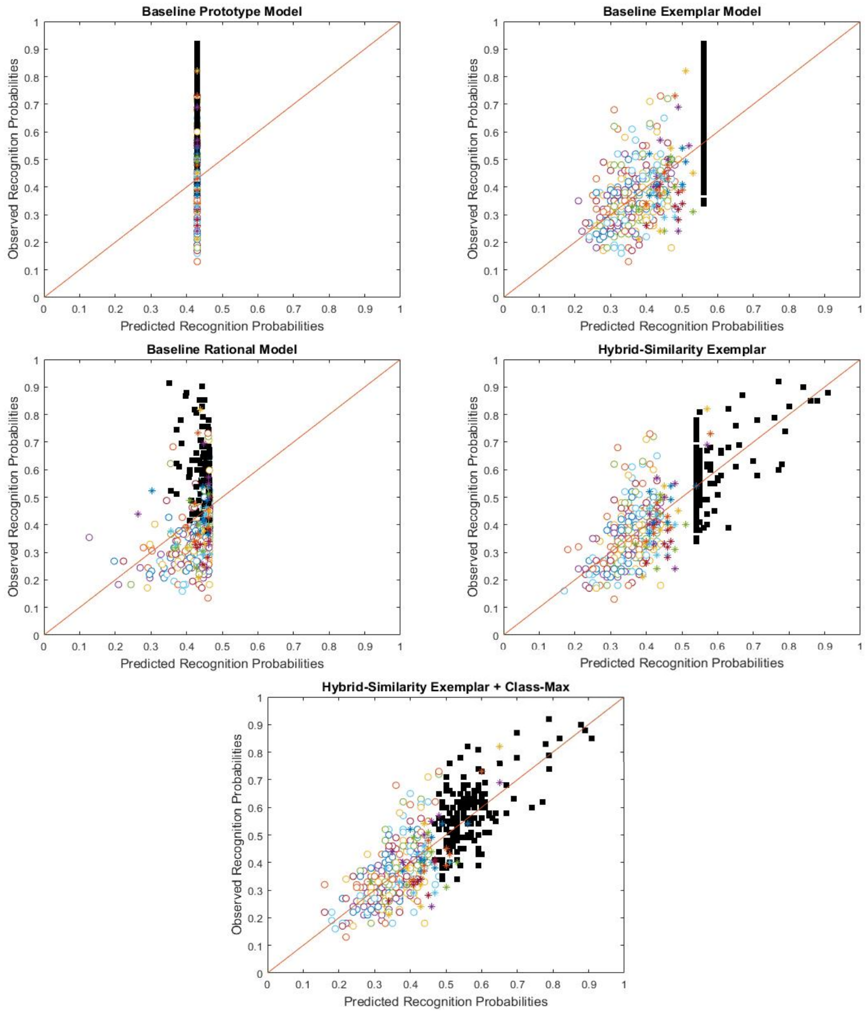
The BIC fits of the models are reported in Table 2. As auxiliary measures, we also report the correlation between the predicted and observed individual-item recognition probabilities and the percentage of variance accounted for. The summary predictions for the four main item types are reported along with the observed data in Table 1. Scatterplots of the observed against predicted old-recognition probabilities for the complete set of 540 individual items are shown in the panels of Figure 3. As can be seen, the prototype model shows a complete failure to account for the old-new recognition data. The rational model accounts for the finding that observers judged the training items and HSNs to be old with higher probability than the standard transfer items, but fails to predict above-chance discrimination between the parents and the HSNs. The exemplar model accounts for observers' overall ability to discriminate among the three main item types. However, as can be seen in Figure 3, the baseline exemplar model shows a complete failure to account for the variability in old-recognition probabilities within the class of old-training items themselves.

Conceptually, the main problem for the baseline exemplar model is that the match of a test item to its own exemplar representation in memory is always equal to one (because the continuous distance between an item and its own exemplar representation is equal to zero). In past work, following ideas

Table 2: Fits of the Models to the Individual-Item Old-New Recognition Data.

Model	# Param	BIC	r	% Var
Prototype	3	60,628.0	0.00	-0.1
Rational	4	60,288.6	0.30	8.9
Exemplar	3	58,677.9	0.70	49.0
Hybrid-Sim Exemplar	6	58,302.8	0.77	59.1
Hybrid-Sim Exemplar + ClassMax	7	58,090.3	0.80	64.7

Figure 3: Predicted vs. Observed Individual-Item Recognition Probabilities. (Solid Squares = Old Training, Open Circles = Standard Transfer, Asterisks = HSN Transfer)



advanced by Lee and Navarro (2002) and Navarro et al. (2003), Nosofsky and Zaki (2003) proposed a hybrid-similarity exemplar model of old-new recognition that combined measures of continuous distance with discrete-feature matching (Tversky, 1977). In this model, matches between test items and exemplars on highly salient discrete features can lead to boosts in the self-match between an item and its own exemplar representation, leading to boosts in overall familiarity. As a proxy for the presence of these types of discrete features in the present stimulus set, an independent group of subjects provided a set of "distinctiveness" ratings, in which they judged the extent to which each item possessed a highly distinctive feature not present in other items in the set. An example of a rock that received a high distinctive-feature rating is shown in Figure 4. Presumably, observers considered the red swirls on the obsidian sample to be highly distinctive, as few rocks in the 540-item set had a feature resembling these swirls.

Let δ_i denote the mean distinctive-feature rating for item i. In brief, in the hybrid-similarity model, the self-similarity between an item i and its own exemplar representation was boosted by the factor $\exp(\beta_1\delta_i)$; the similarity of an HSN transfer item to its parent exemplar was boosted by the factor $\exp(\beta_2\delta_i)$; and the similarity between item i and any other exemplar (i.e. a pair of items unlikely to share a highly salient and distinctive discrete feature) was *reduced* by the factor $\exp(-\beta_3\delta_i)$, where β_1, β_2, and β_3 are freely estimated scaling parameters. The summary fits of the hybrid-similarity exemplar model are reported in Table 2, with a scatterplot of the observed against predicted individual-item recognition probabilities shown in Figure 3. Clearly, the model is moving in the right direction for capturing in better detail the patterns of individual-item recognition probabilities for the class of old training items.

Nevertheless, among the remaining limitations of the model is that there is still a subset of old items that form a

Figure 4: Example of a Rock That Received a High Distinctive-Feature Rating

"vertical wall" in the scatterplot, suggesting that additional factors may be at work in mediating the old-new recognition judgments. We hypothesized that one such factor might be a residual influence of the classification judgments that the observers made on each trial: Although not the central driver of the old-new judgments (see Figure 2), observers may have been somewhat more likely to judge that an item was "old" if they had high confidence in its category membership. To obtain suggestive evidence bearing on this idea, we extended the hybrid-similarity exemplar model by assuming

$$P(\text{Old}|i) = pmix \cdot [F_i^{\gamma} / (F_i^{\gamma} + k)] + (1\text{-}pmix) \cdot [classmax(i)],$$

where *pmix* is a probability-mixing parameter and *classmax*(i) is the "classification confidence" measure described earlier. The summary fits for the mixture model are reported in Table 2, with the revised scatterplot shown in the bottom panel of Figure 3. According to the BIC statistic, the fits are improved, a result corroborated by visual inspection of the scatterplot. Future work will be needed to characterize more precisely the manner in which classification confidence may mediate the old-new recognition judgments.

We should emphasize that these reasonably good quantitative accounts of the individual-item old-new recognition data -- in this complex, real-world, high-dimensional domain -- are still being achieved with a relatively low-parameter model. Even better fits are likely to result by allowing for differential attention-weighting of the component dimensions of the MDS space when observers make their categorization and recognition judgments (Nosofsky, 1991); by making allowance for nonlinear relations between psychological distinctiveness and the direct ratings; and by using improved and more generalizable procedures for detecting and measuring the presence of highly distinctive and salient discrete features. The latter might be achieved, for example, through use of modern deep-learning technology for deriving high-dimensional feature spaces and measures of distinctiveness (e.g., Bhatia & Atka, 2022; Bylinski et al., 2015; Singh et al., 2020). Those directions are crucial ones for future research. In addition, although a full report goes outside the scope of this article, we should note as well that the exemplar model (with its use of a relative-similarity decision rule) yielded outstanding accounts of the categorization transfer data that we observed in our paradigm, replicating past successes reported in other recent studies conducted in this domain (Nosofsky et al., 2018, in press; Sanders & Nosofsky, 2020).

Finally, we should note that it is not obvious to us how the role of matching distinctive features in mediating old-new recognition would be handled in natural fashion within the framework of prototype and clustering models. In exemplar models, major boosts in similarity occur only when matching distinctive features appear in combination with matches along the continuous-dimension values that also compose the test items and the stored exemplars (see Nosofsky & Zaki, 2003, for empirical tests and formal modeling). But in prototype and clustering models, this form of exemplar-specific information – the binding of a distinctive feature to a specific item – is discarded from the memory representations. Thus, it may be difficult to extend prototype and clustering models in natural fashion to allow them to account for the high variability in old-recognition probabilities associated with the old-training items themselves.

Conclusion

The present research provides evidence for the retention of exemplar-specific information in learning categories composed of large numbers of real-world, high-dimensional stimuli. In addition, it is among the first studies to test formal computational models on their ability to predict individual-item old-new recognition judgments – for complex, real-world visual objects – based on the locations of the items in a high-dimensional similarity space. Although the baseline exemplar model outperformed the prototype and the rational-clustering models in accounting for overall patterns of recognition data among the main item types, it failed to account for the extensive variability in recognition probabilities seen within the class of old-training items themselves. We hypothesize that individual old items give rise to different degrees of "self-match" in contacting their own memory representations, and we provided preliminary evidence in favor of this hypothesis through use of an extended hybrid-similarity exemplar model that made reference to an independent set of distinctive-feature ratings for the rocks. However, future research is needed to yield improved formalizations of the discrete-feature matching component of the model and to identify other factors that lead to variability in memory for individual old items.

Acknowledgments

This work was supported in part by NSF grant DUE-1937361.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, *98*(3), 409

Bainbridge, W.A. (2019). Memorability: How what we see influences what we remember. In *Psychology of Learning and Motivation, 70,* 1-27. *Eds., K. Federmeier & D. Beck.*

Bhatia, S., & Aka, A. (2022). Cognitive Modeling With Representations From Large-Scale Digital Data. *Current Directions in Psychological Science*, 09637214211068113.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision research*, *116*, 165-178.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, *91*(1), 1.

Homa, D., Blair, M., McClure, S. M., Medema, J., & Stone, G. (2019). Learning concepts when instances never repeat. *Memory & cognition*, *47*(3), 395-411.

Hooke, R., & Jeeves, T. A. (1961). "Direct Search" Solution of Numerical and Statistical Problems. *Journal of the ACM (JACM)*, *8*(2), 212-229.

Hu, M., & Nosofsky, R. M. (2021). Exemplar-model account of categorization and recognition when training instances never repeat. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*(5140), 1747-1749.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, *9*(1), 43-58.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological review*, *111*(2), 309.

Navarro, D. J., Lee, M. D., Becker, S., Thrun, S., & Obermayer, K. (2003). Combining dimensions and features in similarity-based representations. *Advances in neural information processing systems*, 67-74.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: learning, memory, and cognition*, *14*(4), 700.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of experimental psychology: human perception and performance*, *17*(1), 3.

Nosofsky, R.M. (2017) Exemplar-model accounts of dissociations between categorization and old–new recognition. In J.T. Wixted, J.T. (ed.), *Cognitive Psychology of Memory*, *Vol*. 2 of *Learning and Memory: A Comprehensive Reference, 2nd edition*, Byrne, J.H. (ed.). Oxford: Academic Press.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological review*, *118*(2), 280.

Nosofsky, R. M., Meagher, B., & Kumar, P. (2020). Contrasting Exemplar and Prototype Models in a Natural-Science Category Domain. In *CogSci*.

Nosofsky, R. M., Meagher, B. J., & Kumar, P. (in press). Comparing exemplar and prototype models in a natural-science category domain. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018a). A formal psychological model of classification applied to natural-science category learning. *Current Directions in Psychological Science*, *27*(2), 129-135.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018b). Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. *Journal of Experimental Psychology: General*, *147*(3), 328.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*(2), 530-556.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2020). Search for the missing dimensions: Building a feature-space representation for a natural-science category domain. *Computational Brain & Behavior*, *3*(1), 13-33.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*(4), 247-255.

Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1194.

Osth, A. F., & Dennis, S. (in press). Global matching models of recognition memory. *The Oxford Handbook of Human Memory*.

Sanders, C. A., & Nosofsky, R. M. (2020). Training deep networks to construct a psychological feature space for a natural-object category domain. *Computational Brain & Behavior*, 1-23.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317-1323.

Singh, P., Peterson, J. C., Battleday, R. M., & Griffiths, T. L. (2020). End-to-end deep prototype and exemplar models for predicting human behavior. In Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.

Tversky, A. (1977). Features of similarity. *Psychological review*, *84*(4), 327.